



Mustafa Al Qudah ^{1,2,*}, Ahmad Mohamed ^{1,*} and Syaheerah Lutfi ¹

- ¹ School of Computer Sciences, Universiti Sains Malaysia, Gelugor 11800, Penang, Malaysia
- ² Department of Computer Science, College of Science and Humanities in Al-Sulail,

Prince Sattam bin Abdulaziz University, Kharj 16278, Saudi Arabia

* Correspondence: m.alqudah@psau.edu.sa (M.A.Q.); sufril@usm.my (A.M.)

Abstract: Several studies have been conducted using both visual and thermal facial images to identify human affective states. Despite the advantages of thermal facial images in recognizing spontaneous human affects, few studies have focused on facial occlusion challenges in thermal images, particularly eyeglasses and facial hair occlusion. As a result, three classification models are proposed in this paper to address the problem of thermal occlusion in facial images, with six basic spontaneous emotions being classified. The first proposed model in this paper is based on six main facial regions, including the forehead, tip of the nose, cheeks, mouth, and chin. The second model deconstructs the six main facial regions into multiple subregions to investigate the efficacy of subregions in recognizing the human affective state. The third proposed model in this paper uses selected facial subregions, free of eyeglasses and facial hair (beard, mustaches). Nine statistical features on apex and onset thermal images are implemented. Furthermore, four feature selection techniques with two classification algorithms are proposed for a further investigation. According to the comparative analysis presented in this paper, the results obtained from the three proposed modalities were promising and comparable to those of other studies.

Keywords: affects; emotion; recognition; thermal; occlusion; eyeglass



The human affective state is an important factor that greatly influences our lifestyle, including our thoughts, activities, thinking, focus, and our problem-solving and decision-making abilities. Thus, affective state recognition is of considerable interest to researchers. Affect, mood, and emotion are concepts included in the domain of affective computing. Affect is a general term for a range of feelings that a person can experience. It includes emotions, which are intense feelings that can be directed at a source and are usually short-lived, and moods, which are longer-lasting, less intense, and may not require a specific stimulus [1,2].

Researchers have recently been looking into techniques to recognize human emotions and are applying them to a variety of fields. This includes human–computer interaction to facilitate communication between humans and computers. Human–robot interaction involves understanding how people behave and feel, as this helps the robot to interact with people in an appropriate manner [3]. Human emotion techniques have been used in security applications to identify people who are able to mask their emotions, often referred to as having a "poker face" [4], and in deception detection to detect when someone has not been truthful or accurate in their statements [5,6]. They have also been used for medical applications, including people with autism disorder who may not be able to use their body language, facial expressions, or spoken language to show how they feel. Therefore, they require assistance to help with their specific needs in order to understand and express their emotions [7]. Other medical applications include sleep apnea, which causes a person's breathing to become shallow or to temporarily stop when they are sleeping [8].



Citation: Al Qudah, M.; Mohamed, A.; Lutfi, S. Analysis of Facial Occlusion Challenge in Thermal Images for Human Affective State Recognition. *Sensors* **2023**, *23*, 3513. https://doi.org/10.3390/s23073513

Academic Editor: Cosimo Distante

Received: 13 February 2023 Revised: 11 March 2023 Accepted: 15 March 2023 Published: 27 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). Human affect can be identified by observing the different characteristics of a person, such as their facial expressions, the way they speak, their gestures, the way they stand (their pose), and the direction they are looking. More importantly, previous studies have mentioned that emotional messages are conveyed through facial expressions 55% of the time and through spoken words 7% of the time, while the remaining message is expressed by the manner in which words are spoken, known as paralanguage [9]. Numerous approaches, based on contactless and noninvasive methods, have been applied in emotion recognition and include speech emotion recognition, gesture emotion recognition, and facial emotion recognition, while other studies have concentrated on physiological signals, such as human temperature, cardiac signals, respiration rate, and sweat gland activities.

The human face is often used as a source of information because observing the face is a noninvasive way to collect data. This means that it does not require physical contact or the use of medical instruments. Therefore, it is used in different areas of research, and many studies have been conducted using visible facial images. Despite productive results being reported in the facial emotion recognition literatures, limitations to technique, such as visible images being sensitive to illumination variation [10–13] and diversity in skin color, facial shape, and facial texture, could all contribute to differences in accuracy for the recognition of emotions via facial features. Additionally, different ethical backgrounds and cultural differences could also have an effect, meaning that what is considered happy or sad may be understood differently by different people [10,14]. Furthermore, the discrimination between spontaneous and deceptive emotions is challenging [13]. Therefore, implementing visual-based imaging could contribute to reducing accuracy of affective state recognition [15–17].

Recently, researchers have focused on the thermal imaging approach for affective state recognition because of its ability to reveal a heat pattern consisting of fluctuations in skin temperature caused by involuntary blood circulation that correlates to a particular human expression. Due to the involuntary nature of blood flow, actual emotions may be discernible in the thermal signature [6]. Moreover, the arousal of emotions can cause subtle temperature changes in different regions of the face, such as the forehead, periorbital, supraorbital, tip of the nose, lips, and maxillary [18]. Thermal imaging could be a potential solution to overcoming the illumination variation problem and facilitating the recognition of spontaneous emotion rather than visible images. Although thermal imagery has many tangible advantages, this technique has a variety of drawbacks. One of the limitations is that eyeglasses appear opaque in thermal images and thus decrease the thermal sensitivity of the informative facial regions, such as the inner eye, supraorbital, and periorbital locations [11,19–21]. Moreover, facial hair is another obstacle in thermal imaging which affects the accuracy of temperature measurements [21–23]. A few studies have focused on facial occlusion in affective state recognition. For example, Wang et al. [19] reduced the effects of spectacles by using a thermal temperature pattern; for each frame, the study calculated the mean temperatures of the ambient points, then calculated the temperature differences between each frame and its preceding frame, and finally updated the temperatures of all points for each frame by subtracting the temperature differences. Basu et al. [23] applied a median filter and CLAHE and selected six facial regions, including the forehead, left and right eyes, left and right cheeks, nose, and mouth, proposing Hu's seven moment invariant method to extract features. Nguyen et al. [11] proposed an improved ROI extraction process to overcome eyeglass opacity. Nguyen et al. [20] proposed the fusion of thermal and visual images to overcome the eyeglass occlusion problem.

However, the goal of this study was to investigate more efficient facial subregions in thermal images to overcome eyeglass and facial hair occlusion and to enhance facial affective state recognition. More importantly, the proposed solution in this study comprises three models, with each having its own objective. The first proposed model applies six main ROIs. The eye region is excluded, and other facial areas are employed, including the forehead, tip of the nose, cheeks, mouth, and chin. Moreover, the goal of this model is to recognize the human affective state without implementing important eye ROIs, contrary to previous studies, which means that abandoning eye ROIs could lead to overcoming the eyeglass occlusion challenge. The second proposed model in this study divided the 6 main ROIs into 27 sub-ROIs; the goal of this model is to explore the efficiency of facial sub-ROIs in recognition human affects rather than using the main ROIs, which could have partial occlusion. The third proposed model in this study conducted affective state recognition based on 11 selective sub-ROIs. The selected ROIs located on facial patches with the free existence of facial hair, such as a beard, mustaches, and hair bangs. Therefore, the goal of this model is the recognition of the human affective state even when wearing eyeglasses and with the facial hair occlusion.

2. Previous Studies

This section presents a brief overview of previous studies related to human affective state recognition based on thermal imaging and discusses the main stages, such as dataset collection, preprocessing, facial regions of interest (ROIs), feature extraction methods, and classification algorithms.

2.1. Thermal Dataset

The sophistication of thermal sensors has encouraged researchers to use thermal images to recognize human affects. According to previous studies, the dataset collection stage is an important process; thus, several studies constructed their own dataset [22,24–26], while others used already-published databases, such as KTFE or USTC-NVIE [11,27,28].

Depending on the emotion type, thermal datasets can be classified as posed or spontaneous [29]. Individuals in the posed dataset were asked to express a variety of emotions in order to identify their posed emotional state. This means that this type of emotion does not accurately represent the affective state. In contrast, participants were exposed to stimuli and were unaware that they were being recorded to elicit their spontaneous affective state while their actions were being recorded. As a result, creating spontaneous databases is a difficult process [30].

2.2. Preprocessing

In the preprocessing stage, literatures have proposed various methods to enhance thermal images and localize faces; for instance, HOG and SVM were used by Kopaczka et al. [28] to extract faces from thermal images. In order to identify the region of the initial frame and calculate the head motion, Liu and Yin [31] proposed a model for face identification which comprises a combination of trees with the shard pools of parts drawn from [32]. Otsu thresholding is an approach used in image processing to transform a grayscale image into a black and white image. Wang et al. [13] applied the Otsu thresholding method to create a binary image, and then they analyzed the vertical and horizontal curves of that image to determine the gradient with the highest value; this was used to identify the facial boundary. Latif et al. [33] conducted contrast limited histogram equalization to enhance image contrast. Moreover, to detect the facial region, Mohd et al. [10] used a computer vision technique known as the Viola–Jones boosting algorithm combined with a series of Haar-like features to identify facial regions within a thermal image. Wan et al. [19] proposed a method to help identify the face in an image. The technique uses temperature space, which means that it looks at how hot or cold different parts of an image are. This allows the algorithm to tell which parts are the face and which parts are the background. For face detection, Goulart et al. [24] designed a process to detect a face in a thermal image. This process used three different kinds of filters: median filters, gaussian filters, and a binary filter.

2.3. Region of Interests (ROIs)

Human affects contribute significantly to temperature differences in facial regions. More specifically, the sympathetic nervous system (SNS) responds to human affects by controlling a variety of physiological signals, such as increasing blood flow, which then causes an increase in body temperature, propagated to the surface of the face. Consequently, thermal imaging could detect minor differences in facial temperature [33]. Moreover, variations in facial temperatures are caused by contractions in the facial action unit during human affects [34]. Numerous facial regions have been focused on by previous studies to measure the human affective state, including the forehead, tip of the nose, eyes, mouth, cheeks, and chin [6,10,24,31,33,34]. Differences in temperature values according to a specific emotion type have been reported in numerous studies; for example, Cruz-Albarran et al. [35] mentioned that the temperature of right and left cheeks increased when showing the emotion of being sad or disgusted, whereas the temperature of the maxillary and nose decreased when showing the sad or disgusted emotion. Ioannou et al. [36] showed that forehead temperature decreased when expressing sad or fear emotions and that it increased when showing the anger emotion. Moreover, the study of Jian et al. [37] reported that there is a positive correlation in the cheek and eye regions related to human emotions. Rooj et al. [6] selected facial ROIs such as cheeks, forehead, nose, and maxillary. Kumar et al. [38] proposed facial landmarks with the DenseNet model for facial ROI localization and extraction. Saha et al. [22] proposed the DFTA model to propose eight small facial patches, which contain important information which contributes to differentiating between emotion classes.

2.4. Feature Extraction

The type of selected features plays an important role in classification accuracy; thus, previous studies have used various types of features in their work. For example, statistical features including mean, variance, covariance, median, minimum, maximum, and histogram statistical features [13,20,23–25,28,34,39]. Other feature types which have been used are GLCM features [13,33,40], HOG features (HOG) [28], and LBP features [22,33,34]. Some studies in the literature have also used features from deep learning methods, such as transfer learning from Alex Net [41] and convolutional sparse coding [6].

2.5. Classification

Numerous classification algorithms have been utilized in human affective recognition. For example, the support victor machine algorithm (SVM) has been widely used in previous studies [13,28,31,33,42]. The local discriminant analysis classifier (LDA) [24–26], deep Boltzmann machine (DBN) [13], and deep learning classification algorithms [2,38,41,43,44] have also been utilized.

3. Proposed Methodology

The proposed method in this study consists of three classification models, each with a different number of ROIs. The goal of implementing three models with different numbers of ROIs is to encompass all situations of occlusion presence and explore the efficiency of sub-ROIs in classifying human affects. For example, if a facial image only contains eyeglass occlusion, model one can handle this because it excludes eyes patches from the ROIs. Furthermore, model three can be used to avoid occlusion when facial images include both eyeglasses and facial hair, such as beards or mustaches. The main stages for proposed models are demonstrated in Figure 1. The first stage is the preprocessing of facial images prior to classification; facial images were extracted from their backgrounds and a frontal view process was used to ensure that all the faces had the same correlation due to spontaneous emotions, which are usually accompanied by head movements. Then, 6 main ROIs were cropped from facial patches and subdivided into 27 sub-ROIs. In the proposed models, the following ROIs are used: the first model has 6 main ROIs, the second model has 27 sub-ROIs, and the third model has 11 selective sub-ROIs. The next stage implemented nine statistical features for three models and four types of feature selection algorithms were applied, such as principal component analysis (PCA), analysis of variance (ANOVA), neighborhood components analysis (NCA), and naive Bayes (NB). The last stage



focused on the classification of an affective state based on SVM and MLP for the three proposed models.

Figure 1. Stages of facial affective state recognition using thermal images with occlusion.

3.1. Dataset Selection

This study selected a dataset based on numerous factors. The first one is the recognition of a spontaneous affective state. The second requirement is that onset (beginning of emotion intensity) and apex (maximum emotion intensity) frames should be available for each individual in order to compute statistical features between two images. The third factor focused for the current study is facial occlusion, which includes eyeglasses and facial hair. As a result, the USTC-NVIE [27] database was selected because it satisfied the previous factors. The database provides six basic spontaneous emotions, which are happy, disgust, fear, surprise, anger, and sad. Furthermore, the database applied an evaluation process to ensure the intensity of an emotion for each class through five experienced evaluators. Based on their evaluation report, the study conducted an outlier process to select subjects with a higher emotion intensity, and after detecting the outliers, the number of instances employed for each class was as follows: happy: 99 subjects, disgust: 81 subjects, fear: 55 subjects, surprise: 65 subjects, anger: 56 subjects, and sad: 73 subjects.

3.2. Preprocessing Stage

As demonstrated in Figure 1, facial extraction is the first step in the preprocessing stage. Several approaches have been accomplished in previous studies to extract facial regions, for example, HOG with SVM [45], face detection based on eye coordination and template matching [46], and the Viola–Jones algorithm [33]. The current study applied Goulart et al.'s [24] approach to extract thermal faces by using median and Gaussian filters with further preprocessing stages. More importantly, both onset and apex images have been selected to extract statistical features between them. Therefore, the facial extraction process was applied on both images, which are related to the same subject. Spontaneous emotions were accompanied with facial movements [47]. Therefore, to preserve the same coordination and frontal view for onset and apex images in Figure 2a,b are the onset and apex images before the similarity transformation, and those in Figure 2c,d are the onset and apex images after conducting a similarity transformation.



Figure 2. Image registration process: (**a**,**b**) represents onset and apex thermal images before similarity transformation, respectively; (**c**,**d**) represents onset and apex images after similarity transformation, respectively.

3.3. Selected ROIs

Despite the fact that the aforementioned studies from the literature have focused on several facial regions, such as the forehead, tip of the nose, eyes, mouth, cheeks, and chin [44,48,49], very few previous studies tackled the challenge of manipulating facial occlusion. Therefore, the current study selected facial images with eyeglass occlusion and excluded the important eye region from the selected ROIs. Moreover, this study utilized three classification models to explore the efficiency of ROIs when thermal face comprised occlusion. For example, the first classification model employed six main facial ROIs, including the forehead, tip of the nose, cheeks, mouth, and chin. The second classification model focused on decomposing the previous six main ROIs into 27 sub-ROIs to explore the efficiency of sub-ROIs in the recognition of human affects. The third proposed classification model in this study selected 11 sub-ROIs from 27 sub-ROIs; the selection process conducted based on the criteria of each sub-ROI should be free of facial hair, as the goal of this study is to tackle the challenge of manipulating facial occlusion. Figure 3 demonstrates the proposed three types of selective ROIs for each classification model. The selection of facial ROIs for three proposed classification models in this study is illustrated in the following steps:

- Classification based on six main ROIs: forehead, tip of the nose, left cheek, right cheek, mouth, and chin.
- Classification based on subdividing 6 main ROIs into 27 sub-ROIs as the following:
 - 1. Subdivide forehead main ROI into 12 sub-ROIs.
 - 2. Use tip of the nose ROI without subdividing.
 - 3. Subdivide left cheek main ROI into three sub-ROIs.
 - 4. Subdivide right cheek main ROI into three sub-ROIs.
 - 5. Subdivide mouth main ROI into six sub-ROIs.
 - 6. Subdivide chin main ROI into two sub-ROIs.
- Selected 11 sub-ROIs from 27 sub-ROIs, as demonstrated by the blue color in Figure 3c, labeled as: R9, R10, R11, R12, R13, R14, R17, R22, R23, R24, and R25.
- ROIs extraction in this research accomplished by manually selecting the bounding box on the main facial ROI for the apex image and the coordination of the bounding box applied on onset image to preserve the same ROIs coordination for both the onset and apex images. However, some previous studies have been conducted using the automatic ROI extraction method, but this technique exceeds the scope of the current study.



Figure 3. Selection of ROIs for three proposed classification models: (**a**) represents 6 main ROIs, (**b**) represents 27 sub-ROIs decomposed from 6 main ROIs, (**c**) represents selective 11 sub-ROIs from 27 sub-ROIs.

3.4. Feature Extraction

Before the feature extraction process, the facial regions were converted to temperature values using the equation from [50]. Therefore, the features were calculated based on temperature values instead of gray level intensity values. However, this research proposed nine statistical features to explore the variation in emotion intensity between apex and onset images for each emotion class, as demonstrated in the following:

Equation (1): the mean value of the temperature points (\overline{X}) in the apex image. f_1 represents feature one.

$$f_1 = \overline{X} = \frac{1}{l.m} \sum_{i=1}^{l} \sum_{j=1}^{m} X_{ij}$$
(1)

Equation (2): the mean value of the temperature differences (xd) between the onset and apex images. Variable f_2 represents feature two.

$$f_2 = \overline{xd} = \frac{1}{l.m} \sum_{i=1}^l \sum_{j=1}^m Xd_{i,j}$$
⁽²⁾

Equation (3): variation in the temperature points in the apex image. Variable f_3 represents feature three.

$$f_3 = Variance(x) \tag{3}$$

Equation (4): variation in the temperature differences (Xd) between the onset and apex images. Variable f_4 represents feature four.

$$f_4 = Variance (Xd) \tag{4}$$

Equation (5): the maximum temperature value obtained from the apex image. Variable f_5 represents feature five.

$$f_5 = Max(X) \tag{5}$$

Equation (6): the minimum temperature value obtained from the apex image. Variable f_6 represents feature six.

$$f_6 = Min(X) \tag{6}$$

Equation (7): the mean of the maximum and minimum temperature values obtained from the apex image. Variable f_7 represents feature seven.

$$f_7 = \text{Mean} \left(f_5 + f_6 \right) \tag{7}$$

Equation (8): the median of the temperature values for the apex image. Variable f_8 represents feature eight.

$$f_8 = Median(X) \tag{8}$$

Equation (9): the median of the temperature differences (Xd) between the onset and apex images. Variable f_9 represents feature nine.

$$f_9 = Median(Xd) \tag{9}$$

After conducting statistical features, the number of features for the first, second, and third classification models in this research are 54, 243, and 99 features, respectively. Consequently, for an efficient classification, the feature reduction technique is required due to the large number of features. Four types of feature selection methods were proposed to explore the more efficient features. The selective feature reduction methods include principal component analysis (PCA), analysis of variance (ANOVA), neighborhood components analysis (NCA), and feature selection based on the naïve Bayes (NB) algorithm. After the feature selection algorithms arrange the features according to their importance in the classification, the first 50 features were selected from the ANOVA and PCA algorithms, while first 10 features were selected from the NCA and NB algorithms.

3.5. Classification

For the current study, the support vector machine (SVM), and backpropagation multilayer perceptron (MLP) algorithms were utilized for the classification of six basic emotion classes, including the emotions of happy, disgust, fear, surprise, anger, and sad. To obtain the multiclassification process, the one-against-one technique was used by reducing the multi-classification process to multiple binary classifications between two pairs of classes. Therefore, 15 binary classification models were utilized for each multiclassification. However, to ensure the data of each class were free of noise, the outlier process was performed before the classification by calculating the mean and standard deviation of each class, and instances with more than three standard deviations far from the mean were excluded. More importantly, to preserve balance criteria in binary classification, the down sampling technique was performed. Moreover, a 10-fold cross validation technique was utilized for the validation process. The selected kernel for the SVM classifier was the radial basis function, with an epsilon equal to 0.001, while the configuration of MLP was as follows: the learning rate was equal to 0.01, the number of epochs was 500, the backpropagation rate was 0.2, and the number of hidden layers equaled 6.

4. Experimental Analysis and Discussion

This section demonstrates the experimental results obtained from three classification models and discusses the reported results based on the study's objectives. For each classification model, the process flow was obtained as demonstrated in Figure 4. To evaluate the performance of the proposed models, three statistical analysis methods were obtained, including the precision, F1 score, and Kappa analysis.

4.1. Affective State Recognition Based on Six Main Facial ROIs

The goal of the proposed classification model was to explore the efficiency of six main facial ROIs to classify six basic emotions. Table 1 outlines the mean accuracy results reported from the SVM and MLP classification algorithms with PCA, ANOVA, NCA, and NB feature selectors. As shown in Table 1, the highest mean accuracy results were reported in the sad class, at 98.5%, for SVM-NCA, and the second highest mean accuracy results, at 98.4%, were reported in the sad class for SVM-NB. Moreover, the lowest mean accuracy result was reported to be 62.9% in the fear class from MLP-PCA. Furthermore, Table 2 demonstrates that SVM-NCA reported 95.3%, which is the highest overall mean recognition accuracy result, followed by MLP-NB, which reported 92.1%. The minimum overall mean accuracy results were reported at 73.8% for SVM-PCA. For the evaluation process, Table 3



outlines three statistical evaluation methods, the precision, F1 score, and Kappa, which were reported to be highest from the first two classification algorithms.

Figure 4. Classification process flow with multiple feature selection and classification algorithms.

Table 1. Mean accuracy results for affective state recognition based on six main ROIs reported from the implementation of SVM and MLP classifiers with PCA, ANOVA, NCA, and NB feature selection algorithms.

Classifier/Feature Selector	Happy	Disgust	Fear	Surprise	Anger	Sad	Mean
SVM-PCA	74.6%	72.2%	74.2%	70.2%	71.0%	80.8%	73.8%
SVM-ANOVA	76.7%	82.5%	72.4%	68.6%	77.4%	86.4%	77.3%
SVM-NCA	95.8%	98.2%	90.2%	93.5%	95.3%	98.5%	95.3%
SVM-NB	90.0%	96.4%	84.4%	88.6%	92.5%	98.4%	91.7%
MLP-PCA	84.9%	88.2%	62.9%	76.1%	82.8%	80.2%	79.2%
MLP-ANOVA	75.5%	92.1%	70.2%	75.3%	79.9%	88.9%	80.3%
MLP-NCA	91.5%	97.3%	76.0%	84.1%	90.7%	98.1%	89.6%
MLP-NB	90.9%	97.0%	86.2%	90.5%	90.3%	97.5%	92.1%
Mean	88.1%	94.9%	78.3%	84.7%	88.6%	93.6%	88.0%

Classifier/Feature Selector	Evaluation	Нарру	Disgust	Fear	Surprise	Anger	Sad	Mean
	Precision	0.94	0.96	0.94	0.96	0.93	0.99	0.95
SVM-NCA	F1 score	0.95	0.97	0.92	0.95	0.94	0.99	0.95
	Kappa	0.64	0.64	0.71	0.67	0.70	0.65	0.67
SVM-NB	Precision	0.92	0.90	0.89	0.89	0.92	1.00	0.92
	F1 score	0.91	0.93	0.87	0.89	0.92	0.99	0.92
	Kappa	0.65	0.63	0.72	0.68	0.70	0.65	0.67
	Precision	0.89	0.85	0.93	0.97	0.89	0.90	0.91
MLP-NCA	F1 score	0.90	0.91	0.84	0.90	0.90	0.94	0.90
	Kappa	0.64	0.62	0.74	0.69	0.70	0.64	0.64
MLP-NB	Precision	0.91	0.91	0.97	0.95	0.90	0.92	0.93
	F1 score	0.91	0.94	0.91	0.92	0.90	0.94	0.92
	Kappa	0.65	0.63	0.72	0.68	0.70	0.64	0.67

Table 2. Precision, F1 score, and Kappa statistical evaluation method reported from classification human affective state recognition based on six main ROIs and SVM, MLP classifiers with NCA and NB feature selection algorithms.

Table 3. Mean accuracy results for affective state recognition based on 27 sub-ROIs reported from implementation of SVM and MLP classifiers with PCA, ANOVA, NCA, and NB feature selection algorithms.

Classifier/Feature Selector	Happy	Disgust	Fear	Surprise	Anger	Sad	Mean
SVM-PCA	66.4%	55.5%	74.2%	59.8%	54.1%	59.6%	61.6%
SVM-ANOVA	66.4%	63.3%	69.5%	57.8%	55.6%	78.9%	65.2%
SVM-NCA	93.0%	95.2%	95.6%	94.4%	97.1%	95.3%	95.1%
SVM-NB	94.8%	96.7%	95.3%	93.5%	94.6%	96.0%	95.2%
MLP-PCA	66.7%	76.1%	58.5%	64.7%	82.1%	79.1%	71.2%
MLP-ANOVA	70.0%	66.7%	65.1%	61.1%	69.1%	77.0%	68.2%
MLP-NCA	94.8%	97.3%	91.3%	93.1%	96.8%	96.3%	94.9%
MLP-NB	94.2%	97.6%	93.1%	94.8%	93.2%	96.3%	94.9%
Mean	80.8%	80.7%	80.8%	77.6%	80.1%	84.6%	80.8%

4.2. Affective State Recognition Based on 27 Sub-Facial ROIs

The second classification model in this study explored the efficiency of decomposing 6 main facial ROIs into 27 sub-ROIs with the implementation of 9 statistical features for each sub-ROI. Table 3 outlines the mean accuracy results reported from the proposed method. The highest mean accuracy was reported, at 97.6%, in the disgust class obtained from the MLP-NB classification, and the second highest accuracy result was reported, at 97.3%, in the disgust class from MLP-NCA. The lowest mean accuracy result was reported, at 54.1%, in the anger class from SVM-PCA. The highest overall mean accuracy results were reported to be 95.2% from SVM-NB and 95.1% from SVM-NCA. The overall lowest mean accuracy result reported was 61.6% in SVM-PCA. Table 4 demonstrates statistical evaluation methods related to a higher classification with the feature selection method used in this model.

Classifier/Feature Selector	Evaluation	Нарру	Disgust	Fear	Surprise	Anger	Sad	Overall
	Precision	0.96	0.94	0.95	0.93	0.96	0.96	0.95
SVM-NCA	F1 score	0.94	0.95	0.95	0.94	0.97	0.96	0.95
	Kappa	0.65	0.64	0.70	0.67	0.70	0.66	0.67
SVM-NB	Precision	0.94	0.97	0.92	0.93	0.96	0.98	0.95
	F1 score	0.94	0.97	0.94	0.93	0.96	0.97	0.95
	Kappa	0.65	0.65	0.70	0.67	0.70	0.65	0.67
	Precision	0.93	0.93	0.98	0.95	0.96	0.96	0.95
MLP-NCA	F1 score	0.94	0.95	0.94	0.94	0.96	0.96	0.95
	Kappa	0.64	0.64	0.71	0.68	0.70	0.65	0.67
MLP-NB	Precision	0.93	0.93	0.98	0.95	0.94	0.97	0.95
	F1 score	0.94	0.95	0.95	0.95	0.94	0.97	0.95
	Kappa	0.65	0.64	0.71	0.67	0.70	0.65	0.67

Table 4. Precision, F1 score, and Kappa statistical evaluation method reported from classification human affective state recognition based on 27 sub-ROIs and SVM and MLP classifiers with NCA and NB feature selection algorithms.

4.3. Affective State Recognition Based on Selective 11 Sub-Facial ROIs

The third proposed model for the current study relied on the selection of 11 sub-ROIs from 27 sub-ROIs; the goal of this proposed model was to explore the efficiency of the selected sub-ROIs with free facial hair to avoid thermal occlusion. As demonstrated in Figure 3c, R14 and R17 are the sub-ROIs related to the upper left and right cheeks selected to avoid the existence of hair from a beard. Furthermore, the selected lower sub-ROIs in the forehead are represented by R10 to R12 in Figure 3c, also to avoid patches of forehead that could be covered by hair (bangs). Moreover, only upper and lower lips, represented by R22 to R25, were selected from the mouth region to avoid subregions, which may include facial hair such as mustaches and a beard. However, Table 5 demonstrates the mean accuracy results reported from SVM and MLP classification algorithms and PCA, ANOVA, NCA, and NB feature selectors. The disgust class reported the highest mean recognition accuracy of 96.7% in MLP-NCA and 96.1 in SVM-NCA. Surprisingly, the lowest class was reported to be 42.2% from SVM-ANOVA. Moreover, MLP-NCA reported the highest overall recognition accuracy result of 93.4%, followed by SVM-NB, which reported, at 92.8%, the second highest recognition accuracy. SVM-PCA reported a 51.5% overall lower recognition accuracy. Table 6 demonstrates the statistical evaluation method reported from the classification of the human affective state based on the selected 11 sub-ROIs and SVM and MLP classifiers with NCA and NB feature selection algorithms.

Table 5. Mean accuracy results for affective state recognition based on selective 11 sub-ROIs reported from implementation of SVM and MLP classifiers with PCA, ANOVA, NCA, and NB feature selection algorithms.

Classifier/Feature Selector	Happy	Disgust	Fear	Surprise	Anger	Sad	Mean
SVM-PCA	56.4%	49.4%	48.4%	55.9%	47.0%	51.9%	51.5%
SVM-ANOVA	63.0%	55.9%	61.6%	42.2%	55.9%	79.2%	59.6%
SVM-NCA	90.9%	96.1%	90.2%	90.5%	94.6%	94.1%	92.7%
SVM-NB	91.5%	92.7%	93.5%	91.2%	93.2%	94.4%	92.8%
MLP-PCA	48.8%	50.6%	50.5%	52.9%	55.9%	50.6%	51.6%
MLP-ANOVA	68.8%	67.9%	62.8%	59.5%	59.9%	86.0%	67.5%
MLP-NCA	93.6%	96.7%	90.9%	92.2%	91.7%	95.3%	93.4%
MLP-NB	92.4%	94.2%	91.6%	93.5%	91.8%	93.5%	92.8%
Mean	75.7%	75.4%	73.7%	72.2%	73.7%	80.6%	75.2%

Classifier/Feature Selector	Evaluation	Happy	Disgust	Fear	Surprise	Anger	Sad	Overall
	Precision	0.93	0.92	0.95	0.91	0.92	0.94	0.93
SVM-NCA	F1 score	0.92	0.94	0.92	0.91	0.93	0.94	0.93
	Kappa	0.65	0.64	0.71	0.67	0.70	0.65	0.67
SVM-NB	Precision	0.93	0.92	0.93	0.92	0.94	0.93	0.93
	F1 score	0.92	0.92	0.93	0.92	0.93	0.94	0.93
	Kappa	0.65	0.65	0.70	0.67	0.70	0.65	0.67
	Precision	0.93	0.93	1	0.93	0.91	0.92	0.94
MLP-NCA	F1 score	0.93	0.95	0.95	0.93	0.91	0.94	0.94
	Kappa	0.65	0.64	0.72	0.67	0.70	0.65	0.67
MLP-NB	Precision	0.93	0.92	0.96	0.91	0.95	0.92	0.93
	F1 score	0.93	0.93	0.94	0.92	0.93	0.93	0.93
	Kappa	0.65	0.64	0.71	0.67	0.71	0.65	0.67

Table 6. Precision, F1 score, and Kappa statistical evaluation method reported from classification human affective state recognition based on selective 11 sub-ROIs and SVM and MLP classifiers with NCA and NB feature selection algorithms.

4.4. Comparative Study

The highest results from the previous proposed classification models have been reported by two feature selection algorithms, namely, NCA and NB, while the lowest outcomes were identified by PCA and ANOVA. This finding may refer to the significance of feature selection algorithms in the classification process. Moreover, as aforementioned, the number of selected features could play a significant role in the classification results. In the current study, 50 higher ranked features were opted from PCA and ANOVA, while 10 higher ranked features were selected from NCA and NB. The reason for selecting 50 features from PCA and ANOVA relied on experimental trails to achieve higher accuracy results. More importantly, Figure 5 documents a comparison between the results reported from three proposed models. As shown in Figure 5, the highest overall mean accuracy result in model one was 95.3%, reported from the SVM-NCA classifier, and the highest overall mean accuracy result in model two was 95.2%, reported from SVM-NB classifiers, while the highest overall mean accuracy result in model three was 93.4%, reported from MLP-NCA. However, the results from the three proposed models appeared to be nearly identical, which means that the proposed method of decomposing main facial patches into sub-regions could help the researcher to avoid occluded facial regions in thermal images, such as eyeglasses and facial hair. Furthermore, the results of this study show that, despite using a small number of facial ROIs, the classification performance is still promising. More importantly, the current study's findings show that increasing the number of features does not always result in a higher accuracy, while few robust features could report more significant impact.

After a comparison has been made between three proposed models, it is important to compare the proposed models with those conducted in the literature. Therefore, the current study selected models from other studies which selected their data from the USTC-NVIE [27] database. Table 7 demonstrates the comparative results obtained from the current study with other methods from the literature based on the UTSC-NVIE database. As shown in Table 7, the results reported from the current study outperform the results reported from other studies, except for the study in [41], which reported higher mean accuracy results. The reason for this could refer to the type of features that were used.



Figure 5. Comparison between three proposed classification models.

Methods	Features	ROIs	Classifier	Accuracy	Precession	F1 Score	Kappa
[22]	LBP	Selected facial sub-ROIs	Multiclass SVM	78.16%	78.3%	0.774	-
[51]	Statistical features	100 sub-ROIs without facial area	SVM	76.45%	76.6%	0.753	0.637
[41]	Transfer Learning	Hybrid approach	Multiclass SVM	99.3%	98.5%	0.984	0.93
[6]	Sparce coded filter	Selected facial sub-ROIs	SVM	66.1%	67.1%	0.663	0.179
	Statistical features	6 main facial ROIs	SVM-NCA	95.3%	95%	0.95	0.67
Our Proposed methods	Statistical features	27 facial sub-ROIs	SVM-NB	95.2%	95%	0.95	0.67
	Statistical features	Selective 11 facial sub-ROIs	SVM-NB	92.8%	93%	0.93	0.67

Table 7. Comparative results with other literature studies based on USTC-NVIE database.

This study also validated its findings by comparing them to previous studies based on visual facial images. Table 8 outlines the comparison based on visual images and shows that the mean accuracy results reported from the current study are competitive with other results.

Methods	Datasets	Features	Classifier	Accuracy
Proposed model one	USTC-NVIE	Statistical features	SVM-NCA	95.3%
Proposed model two	USTC-NVIE	Statistical features	SVM-NB	95.2%
Proposed model three	USTC-NVIE	Statistical features	SVM-NB	92.8
[52]	JAFFE	Haar, Gabor, LBP	SVM, KNN	89.5%
[53]	JAFFE	facial landmarks, and center of gravity	SVM	91.9%
[54]	JAFFE	DBN	ANN	90.95%
[55]	KDEF	Transfer learning	CNN	98.78%
	JAFFE	Transfer learning	CNN	99.52%

Table 8. Comparative results with other studies in the literature based on visual facial images dataset.

5. Conclusions

This paper proposed three modalities for spontaneous affective state recognition based on facial thermal images with eyeglass and facial hair occlusion. The main objective of this paper was to explore the facial sub-regions which were free of eyeglass and hair occlusion and more efficient in exploring the human affective state. The three proposed models were dependent on each other. The first model focused on the classification of affective states based on six main facial ROIs, and the eyeglass location was excluded. The second model decomposed the 6 main ROIs into 27 sub-ROIs to explore the efficiency of sub-facial regions in the classification of affective states. The third model in this paper selected 11 sub-ROIs from 27 ROIs to explore the ability of avoiding facial hair regions in thermal images. In comparison to previous studies, the results reported from the three proposed models demonstrate a higher mean accuracy: 95.3%, 95.2%, and 92.8% for models one, two, and three, respectively. Furthermore, the results of this study show the importance of feature selection techniques in improving classification accuracy. However, in future studies, we will focus on the automatic extraction of ROIs and employ deep learning algorithms to improve the recognition of human affective states.

Author Contributions: Conceptualization, M.A.Q., A.M. and S.L.; methodology. M.A.Q.; software, M.A.Q.; validation, M.A.Q., A.M. and S.L.; formal analysis, M.A.Q., A.M. and S.L.; investigation, M.A.Q.; resources, M.A.Q.; data curation, M.A.Q.; writing original draft preparation, M.A.Q.; supervision, A.M. and S.L.; project administration, A.M.; funding acquisition, A.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Ministry of Higher Education Malaysia for Fundamental Research Grant Scheme with project code: FRGS/1/2020/STG07/USM/02/12.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing not applicable.

Conflicts of Interest: The authors declare no conflict of interest regarding the publication of this paper.

References

- 1. Shouse, E. Feeling, Emotion, Affect. M/C J. 2005, 8. [CrossRef]
- Nayak, S.; Panda, S.K.; Uttarkabat, S. A Non-contact Framework based on Thermal and Visual Imaging for Classification of Affective States during HCI. In Proceedings of the 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 15–17 June 2020; pp. 653–660. [CrossRef]
- Desideri, L.; Ottaviani, C.; Malavasi, M.; di Marzio, R.; Bonifacci, P. Emotional processes in human-robot interaction during brief cognitive testing. *Comput. Human Behav.* 2019, 90, 331–342. [CrossRef]
- Lee, M.S.; Cho, Y.R.; Lee, Y.K.; Pae, D.S.; Lim, M.T.; Kang, T.K. PPG and EMG based emotion recognition using convolutional neural network. In *Proceedings of the 16th International Conference on Informatics in Control, Automation and Robotics, Prague, Czech Republic, 29 July 2019*; Gusikhin, O., Madani, K., Zaytoon, J., Eds.; SciTePress: Prague, Czech Republic, 2019; pp. 595–600. [CrossRef]

- 5. Nancarrow, A.F.; Gilpin, A.T.; Thibodeau, R.B.; Farrell, C.B. Knowing what others know: Linking deception detection, emotion knowledge, and Theory of Mind in preschool. *Infant Child Dev.* **2018**, 27, e2097. [CrossRef]
- Rooj, S.; Antesh, U.; Bhattacharya, S.; Routray, A.; Mandal, M.K. Emotion Classification of Facial Thermal Images using Sparse Coded Filters. In Proceedings of the IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society, Singapore, 18–21 October 2020; pp. 453–458. [CrossRef]
- Chen, C.-H.; Lee, I.J.; Lin, L.-Y. Augmented reality-based self-facial modeling to promote the emotional expression and social skills of adolescents with autism spectrum disorders. *Res. Dev. Disabil.* 2015, 36, 396–403. [CrossRef]
- Usman, M.; Evans, R.; Saatchi, R.; Kingshott, R.; Elphick, H. Non-invasive respiration monitoring by thermal imaging to detect sleep apnoea. In Proceedings of the 32nd International Congress and Exhibition on Condition Monitoring and Diagnostic Engineering Management, Huddersfield, UK, 3–5 September 2019.
- 9. Huang, Y.; Chen, F.; Lv, S.; Wang, X. Facial expression recognition: A survey. Symmetry 2019, 11, 1189. [CrossRef]
- 10. Mohd, M.N.H.; Kashima, M.; Sato, K.; Watanabe, M. Mental stress recognition based on non-invasive and non-contact measurement from stereo thermal and visible sensors. *Int. J. Affect. Eng.* **2015**, *14*, 9–17. [CrossRef]
- Nguyen, T.; Tran, K.; Nguyen, H. Towards thermal region of interest for human emotion estimation. In Proceedings of the 2018 10th International Conference on Knowledge and Systems Engineering (KSE), Ho Chi Minh City, Vietnam, 13 December 2018; pp. 152–157. [CrossRef]
- 12. Wang, S.; Pan, B.; Chen, H.; Ji, Q. Thermal Augmented Expression Recognition. IEEE Trans Cybern 2018, 48, 2203–2214. [CrossRef]
- 13. Wang, S.; He, M.; Gao, Z.; He, S.; Ji, Q. Emotion recognition from thermal infrared images using deep Boltzmann machine. *Front. Comput. Sci.* **2014**, *8*, 609–618. [CrossRef]
- 14. Yan, X.; Andrews, T.J.; Jenkins, R.; Young, A.W. Cross-cultural differences and similarities underlying other-race effects for facial identity and expression. *Q. J. Exp. Psychol.* **2016**, *62*. [CrossRef]
- 15. Wang, S.; Liu, Z.; Wang, Z.; Wu, G.; Shen, P.; He, S.; Wang, X. Analyses of a Multimodal Spontaneous Facial Expression Database. *IEEE Trans. Affect. Comput.* **2013**, *4*, 34–46. [CrossRef]
- Bhowmik, M.K.; Saha, P.; Singha, A.; Bhattacharjee, D.; Dutta, P. Enhancement of robustness of face recognition system through reduced gaussianity in Log-ICA. *Expert Syst. Appl.* 2019, *116*, 96–107. [CrossRef]
- 17. Zhang, X.; Yin, L.; Cohn, J.F.; Canavan, S.; Reale, M.; Horowitz, A.; Liu, P.; Girard, J.M. Bp4d-spontaneous: A high-resolution spontaneous 3d dynamic facial expression database. *Image Vis. Comput.* **2014**, *32*, 692–706. [CrossRef]
- Al Qudah, M.M.; Mohamed, A.S.; Lutfi, S.L. Affective State Recognition Using Thermal-Based Imaging: A Survey. Comput. Syst. Sci. Eng. 2021, 37, 47–62. [CrossRef]
- Wang, Y.; Jiang, X.; Yang, M.; Zhang, D.; Yi, X.; Nguyen, H.; Kotani, K.; Chen, F.; Le, B. Estimation of human emotions using thermal facial information. In Proceedings of the 5th International Conference on Graphic and Image Processing (ICGIP 2013), Hong Kong, China, 10 January 2014. [CrossRef]
- Nguyen, H.; Chen, F.; Kotani, K.; Le, B. Fusion of visible images and thermal image sequences for automated facial emotion estimation. J. Mobile Multimed. 2014, 10, 294–308.
- Abd Latif, M.; Yusof, H.M.; Sidek, S.; Rusli, N. Thermal imaging based affective state recognition. In Proceedings of the 2015 IEEE International Symposium on Robotics and Intelligent Sensors (IRIS), Langkawi, Malaysia, 18 October 2015; pp. 214–219. [CrossRef]
- 22. Saha, P.; Bhattacharjee, D.; De, B.K.; Nasipuri, M. A Thermal Blended Facial Expression Analysis and Recognition System Using Deformed Thermal Facial Areas. *Int. J. Image Graph.* 2022, 22, 2250049. [CrossRef]
- Basu, A.; Routray, A.; Shit, S.; Deb, A.K. Human emotion recognition from facial thermal image based on fused statistical feature and multi-class SVM. In Proceedings of the 2015 Annual IEEE India Conference (INDICON), New Delhi, India, 17 December 2015; pp. 1–5. [CrossRef]
- Goulart, C.; Valadão, C.; Delisle-Rodriguez, D.; Caldeira, E.; Bastos, T. Emotion analysis in children through facial emissivity of infrared thermal imaging. *PLoS ONE* 2019, 14, e0212928. [CrossRef] [PubMed]
- Khan, M.M.; Ward, R.D.; Ingleby, M. Classifying pretended and evoked facial expressions of positive and negative affective states using infrared measurement of skin temperature. ACM Trans. Appl. Percept. 2009, 6, 1–22. [CrossRef]
- 26. Cross, C.; Skipper, J.; Petkie, D. Thermal imaging to detect physiological indicators of stress in humans. In Proceedings of the SPIE Defense, Security, and Sensing, Baltimore, MD, USA, 22 May 2013. [CrossRef]
- Wang, S.; Liu, Z.; Lv, S.; Lv, Y.; Wu, G.; Peng, P.; Chen, F.; Wang, X. A Natural Visible and Infrared Facial Expression Database for Expression Recognition and Emotion Inference. *IEEE Trans. Multimed.* 2010, 12, 682–691. [CrossRef]
- Kopaczka, M.; Kolk, R.; Merhof, D. A fully annotated thermal face database and its application for thermal facial expression recognition. In Proceedings of the 2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Houston, TX, USA, 12 July 2018; pp. 1–6. [CrossRef]
- 29. Haamer, R.E.; Rusadze, E.; Lsi, I.; Ahmed, T.; Escalera, S.; Anbarjafari, G. Review on emotion recognition databases. *Hum. Robot Interact. Theor. Appl* **2017**, *3*, 39–63. [CrossRef]
- 30. Pantic, M.; Valstar, M.; Rademaker, R.; Maat, L. Web-based database for facial expression analysis. In Proceedings of the 2005 IEEE International Conference on Multimedia and Expo, Amsterdam, The Netherlands, 24 October 2005; p. 5. [CrossRef]

- Liu, P.; Yin, L. Spontaneous facial expression analysis based on temperature changes and head motions. In Proceedings of the 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Ljubljana, Slovenia, 23 July 2015; pp. 1–6. [CrossRef]
- Zhu, X.; Ramanan, D. Face detection, pose estimation, and landmark localization in the wild. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 2879–2886.
- Latif, M.; Yusof, M.H.; Sidek, S.; Rusli, N. Texture descriptors based affective states recognition-frontal face thermal image. In Proceedings of the 2016 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES), Kuala Lumpur, Malaysia, 4–8 December 2016; pp. 80–85. [CrossRef]
- Carrapiço, R.; Mourao, A.; Magalhaes, J.; Cavaco, S. A comparison of thermal image descriptors for face analysis. In Proceedings of the 2015 23rd European Signal Processing Conference (EUSIPCO), Nice, France, 31 August 2015; pp. 829–833. [CrossRef]
- 35. Cruz-Albarran, I.A.; Benitez-Rangel, J.P.; Osornio-Rios, R.A.; Morales-Hernandez, L.A. Human emotions detection based on a smart-thermal system of thermographic images. *Infrared Phys. Technol.* **2017**, *81*, 250–261. [CrossRef]
- 36. Ioannou, S.; Gallese, V.; Merla, A. Thermal infrared imaging in psychophysiology: Potentialities and limits. *Psychophysiology* **2014**, *51*, 951–963. [CrossRef] [PubMed]
- Jian, B.-L.; Chen, C.-L.; Huang, M.-W.; Yau, H.-T. Emotion-Specific Facial Activation Maps Based on Infrared Thermal Image Sequences. *IEEE Access* 2019, 7, 48046–48052. [CrossRef]
- Kyal, C.K.; Poddar, H.; Reza, M. Human Emotion Recognition from Spontaneous Thermal Image Sequence Using GPU Accelerated Emotion Landmark Localization and Parallel Deep Emotion Net. In Proceedings of the International Conference on Innovative Computing and Communications: Proceedings of ICICC 2020, Delhi, India, 21–23 February 2020; pp. 931–943. [CrossRef]
- Pérez-Rosas, V.; Narvaez, A.; Burzo, M.; Mihalcea, R. Thermal imaging for affect detection. In Proceedings of the 6th International Conference on PErvasive Technologies Related to Assistive Environments—PETRA '13, Rhodes, Greece, 29–31 May 2013; pp. 1–4. [CrossRef]
- 40. Latif, M.; Sidek, S.; Rusli, N.; Fatai, S. Emotion detection from thermal facial imprint based on GLCM features. *ARPN J. Eng. Appl. Sci.* 2016, *11*, 345–350.
- Shaees, S.; Naeem, H.; Arslan, M.; Naeem, M.R.; Ali, S.H.; Aldabbas, H. Facial emotion recognition using transfer learning. In Proceedings of the 2020 International Conference on Computing and Information Technology (ICCIT-1441), Tabuk, Saudi Arabia, 9–10 September 2020; pp. 1–5. [CrossRef]
- Boccanfuso, L.; Wang, Q.; Leite, I.; Li, B.; Torres, C.; Chen, L.; Salomons, N.; Foster, C.; Barney, E.; Ahn, Y.A. A thermal emotion classifier for improved human-robot interaction. In Proceedings of the 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), New York, NY, USA, 26–31 August 2016; pp. 718–723. [CrossRef]
- Liu, Y.; Chen, X.; Wang, Z.; Wang, Z.J.; Ward, R.K.; Wang, X. Deep learning for pixel-level image fusion: Recent advances and future prospects. *Inf. Fusion* 2018, 42, 158–173. [CrossRef]
- 44. Nayak, S.; Nagesh, B.; Routray, A.; Sarma, M. A Human–Computer Interaction framework for emotion recognition through time-series thermal video sequences. *Comput. Electr. Eng.* **2021**, *93*, 107280. [CrossRef]
- 45. Kopaczka, M.; Schock, J.; Nestler, J.; Kielholz, K.; Merhof, D. A combined modular system for face detection, head pose estimation, face tracking and emotion recognition in thermal infrared images. In Proceedings of the 2018 IEEE International Conference on Imaging Systems and Techniques (IST), Krakow, Poland, 16–18 October 2018; pp. 1–6. [CrossRef]
- Sharma, N.; Dhall, A.; Gedeon, T.; Goecke, R. Modeling Stress Using Thermal Facial Patterns: A Spatio-temporal Approach. In Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, Geneva, Switzerland, 2–5 September 2013; pp. 387–392. [CrossRef]
- 47. Bian, C.; Zhang, Y.; Yang, F.; Bi, W.; Lu, W. Spontaneous facial expression database for academic emotion inference in online learning. *IET Comput. Vis.* **2019**, *13*, 329–337. [CrossRef]
- Murtaza, M.; Sharif, M.; AbdullahYasmin, M.; Ahmad, T. Facial expression detection using six facial expressions hexagon (SFEH) model. In Proceedings of the 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 7–9 January 2019; pp. 0190–0195. [CrossRef]
- Saha, P.; Bhattacharjee, D.; De, B.K.; Nasipuri, M. Characterization and recognition of mixed emotional expressions in thermal face image. In Proceedings of the Infrared Imaging Systems: Design, Analysis, Modeling, and Testing xxvii, Baltimore, MD, USA, 17–21 April 2016; pp. 184–193. [CrossRef]
- 50. Ng, E.; Yee, G.C.; Hua, T.J.; Kagathi, M. Analysis of normal human eye with different age groups using infrared images. *J. Med. Syst.* 2009, *33*, 207–213. [CrossRef]
- 51. Shi, X.; Wang, S.; Zhu, Y. Expression recognition from visible images with the help of thermal images. In Proceedings of the 5th ACM on International Conference on Multimedia Retrieval, Ottawa, ON, Canada, 23–26 June 2015; pp. 563–566. [CrossRef]
- Liew, C.F.; Yairi, T. Facial expression recognition and analysis: A comparison study of feature descriptors. *IPSJ Trans. Comput. Vis. Appl.* 2015, 7, 104–120. [CrossRef]
- Alshamsi, H.; Meng, H.; Li, M. Real time facial expression recognition app development on mobile phones. In Proceedings of the 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Changsha, China, 13–15 August 2016; pp. 1750–1755. [CrossRef]

- 54. Zhao, X.; Shi, X.; Zhang, S. Facial expression recognition via deep learning. *IETE Tech. Rev.* 2015, *32*, 347–355. [CrossRef]
- 55. Akhand, M.A.H.; Roy, S.; Siddique, N.; Kamal, M.A.S.; Shimamura, T. Facial Emotion Recognition Using Transfer Learning in the Deep CNN. *Electronics* **2021**, *10*, 1036. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.